

Claims

1. A web content converting system for converting a large display screen web document into a small display screen web document, the system comprising:

a preprocessor for standardizing a non-standard web document having an erroneous tag to output the standardized web document in a data format suitable for analysis;

a client profile analyzer for extracting and managing client information;

a structure analyzer for receiving the web document standardized in the preprocessor to set the web document to a content unit piece (component) according to a document analysis algorithm;

an image converter for extracting information on an image encoding/decoding procedure and an image size included in the web document;

a component block extractor for grouping the set content unit piece (component) to similar groups within a range not exceeding a maximal width by using an attribution value of the content unit piece (component) and client performance information;

a component block categorizer for categorizing each of component blocks generated by the component block extractor into index and body content portions in accordance with a content characteristic;

an index generator for extracting information on image or text index from the component block categorized into the index portion, and generating a script file and an additional tag collection for expressing the extracted information;

a voice markup generator for converting a text-centered body content block into a voice markup language to perform a voice supporting function; and

a HyperText Markup Language (HTML) generator for rearranging and reconstructing the generated content object elements according to a document pattern to generate the small display screen web document.

2. The web content converting system of claim 1, wherein the web content converting system is installed at any one of three layers of a web server, a client and a proxy.

3. A web content converting method for converting a large display screen web document into a small display screen web document, the method comprising:

5 a preprocessing step for standardizing a non-standard web document including an erroneous tag to output the standardized web document in a data format suitable for analysis;

a web document analyzing step for receiving the standardized web document and analyzing a tag according to a document analysis algorithm to set the web document to a content unit piece (component);

10 a component block setting step for grouping the set content unit piece (component) to similar groups within a range not exceeding a maximal width by using an attribution value of the content unit piece (component) and client performance information;

15 a component block categorizing step for categorizing each of component blocks generated by the component block extractor into index and body content portions in accordance with a content characteristic;

an index generating step for extracting information on image or text index from the component block categorized into the index portion, and generating a script file and an additional tag collection for expressing the extracted information;

20 a voice-markup generating step for converting a text-centered body content block into a voice markup language to perform a voice supporting function; and

a HyperText Markup Language (HTML) generating step for rearranging and reconstructing the generated content object elements according to a document pattern to generate the small display screen web document.

25 4. The web content converting method of claim 3, wherein in the web document analyzing step, a tag such as <TABLE>, <TR>, <TD>, , etc. is mainly analyzed, and a specific <TD> tag is defined as a component to be used as a minimal unit for the content unit analysis.

30 5. The web content converting method of claim 3, wherein in the component block setting step, a component tree is inputted to check initial width information for all component nodes, and it is checked whether or not a sibling node of a current component node exists, and if existing, similar sibling nodes are bundled and grouped within the range not exceeding the maximal width (MAX_WIDTH).

6. The web content converting method of claim 3, wherein the component block categorizing step comprises the steps of:

5 receiving a component block tree to visit all component blocks while to compare a content pattern of the component block;

determining an index type if a resultant value of the pattern comparison exceeds a certain critical value;

10 setting a type of the index-determined block to each of an image index (INDEX_I) or a text index (INDEX_T) depending on whether a data type of the content is an image or a text; and

15 categorizing the block not being the index into the body, and categorizing the voice body (BODY_V) for performing the converting into the voice supporting document and the general body (BODY_G) processed as other general content blocks.